Hindawi Publishing Corporation
EURASIP Journal on Bioinformatics and Systems Biology
Volume 2010, Article ID 947564, 10 pages
doi:10.1155/2010/947564

*Research Article*

# A Hypothesis Test for Equality of Bayesian Network Models

**Anthony Almudevar**

*Department of Computational Biology, University of Rochester, 601 Elmwood Avenue, Rochester, NY 14642, USA*

Correspondence should be addressed to Anthony Almudevar, anthony_almudevar@urmc.rochester.edu

Received 26 March 2010; Revised 9 July 2010; Accepted 5 August 2010

Academic Editor: A. Datta



Bayesian network models are commonly used to model gene expression data. Some applications require a comparison of the network structure of a set of genes between varying phenotypes. In principle, separately fit models can be directly compared, but it is difficult to assign statistical significance to any observed differences. There would therefore be an advantage to the development of a rigorous hypothesis test for homogeneity of network structure. In this paper, a generalized likelihood ratio test based on Bayesian network models is developed, with significance level estimated using permutation replications. In order to be computationally feasible, a number of algorithms are introduced. First, a method for approximating multivariate distributions due to Chow and Liu (1968) is adapted, permitting the polynomial-time calculation of a maximum likelihood Bayesian network with maximum indegree of one. Second, sequential testing principles are applied to the permutation test, allowing significant reduction of computation time while preserving reported error rates used in multiple testing. The method is applied to gene-set analysis, using two sets of experimental data, and some advantage to a pathway modelling approach to this problem is reported.

## 1. Introduction

Graphical models play a central role in modelling genomic data, largely because the pathway structure governing the interactions of cellular components induces statistical dependence naturally described by directed or undirected graphs [1–3]. These models vary in their formal structure. While a *Boolean network* can be interpreted as a set of state transition rules, *Bayesian* or *Markov networks* reduce to static multivariate densities on random vectors extracted from genomic data. Such densities are designed to model coexpression patterns resulting from functional cooperation. Our concern will be with this type of multivariate model. Although the ideas presented here extend naturally to various forms of genomic data, to fix ideas we will refer specifically to multivariate samples of microarray gene expression data.

In this paper, we consider the problem of comparing network models for a common set of genes under varying phenotypes. In principle, separately fit models can be directly compared. This approach is discussed in [3] and is based on distances definable on a space of graphs. Significance levels are estimated using replications of random graphs similar in structure to the estimated models.

The algorithm proposed below differs significantly from the direct graph approach. We will formulate the problem as a two-sample test in which significance levels are estimated by randomly permuting phenotypes. This requires only the minimal assumption of independence with respect to subjects.

Our strategy will be to confine attention to Bayesian network models (Section 2). Fitting Bayesian networks is computationally difficult, so a simplified model is developed for which a polynomial-time algorithm exists for maximum likelihood calculations. A two-sample hypotheses test based on the general likelihood ratio test statistic is introduced in Section 3. In Section 4, we discuss the application of sequential testing principles to permutation replications. This may be done in a way which permits the reporting of error rates commonly used in multiple testing procedures. In Section 5, the methodology is applied to the problem of *gene set* (GS) analysis, in which high dimensional arrays of gene expression data are screened for *differential expression* (DE) by comparing gene sets defined by known functional relationships,

in place of individual gene expressions. This follows the paradigm originally proposed in *gene set enrichment analysis* (GSEA) [4–6]. The method will be applied to two well-known microarray data sets.

An R library of source code implementing the algorithms proposed here may be downloaded at http://www.urmc.rochester.edu/biostat/people/faculty/almudevar.cfm.

## 2. Network Models

A graphical model is developed by defining each of $n$ genes as a graph node, labelled by gene expression level $X_i$ for gene $i$. The model incorporates two elements, first, a *topology* $G$ (a directed or undirected graph on the $n$ nodes), then, a multivariate distribution $f$ for $X = (X_1, \ldots, X_n)$ which conforms to $G$ in some well defined sense. In a *Bayesian network* (BN), model $G$ is a *directed acyclic graph* (DAG), and $f$ assumes the form

$$f(x) = \prod_{i=1}^{n} f_i\left(x_i \mid x_j, j \in Pa_G(i)\right), \tag{1}$$

where $Pa_G(i)$ is the set of parents of node $i$. Intuitively, $f_i(x_i|x_j, j \in Pa_G(i))$ describes a causal relationship between node $i$ and nodes $Pa_G(i)$.

The advantage of (1) is the reduction in the degrees of freedom of the model while preserving coexpression structure. Also, some flexibility is available with respect to the choice of the conditional densities of (1), with Gaussian, multinomial, and Gamma forms commonly used [7]. We note that BNs are commonly used in many genomic applications [7–9].

*2.1. Gaussian Bayesian Network Model.* For this application, we will use the Gaussian BN. These models are naturally expressed using a linear regression model of node $i$ data $X_i$ on the data $X_j$, $j \in Pa_G(i)$. In [10], it is noted that in microarray data gene expression levels are aggregated over large numbers of individual cells. Linear correlations are preserved under this process, but other forms of dependence generally will not be, so we can expect linear regression to capture the dominant forms of interaction which are statistically observable. In this case the maximum log-likelihood function for a given topology reduces to

$$L(G) = \sum_i -\ln(\mathrm{MSE}[Pa_G(i)]), \tag{2}$$

where $\mathrm{MSE}[Pa_G(i)]$ is the mean squared error of a linear regression fit of the offspring expressions onto those of the parents.

*2.2. Restricted Bayesian Networks.* Fitting BNs involves optimization over the space of topologies and hence is computationally intensive [9]. While exact algorithms are available [11], they will generally require too great a computation time for the application described below. A recent application of exact techniques to the problem of pedigree reconstruction (a BN with maximum indegree of 2) was described in [12]. Using methods proposed in [13] the exact computation of the maximum likelihood of a pedigree with 29 individuals (nodes) required 8 minutes. The author of [12] agrees with the conclusion reported in [13], that the method is not viable for BNs with greater than 32 nodes.

It is possible to control the size of the computation by placing a cap $K$ on the permissable indegree of each node, though the problem remains difficult even for $K = 2$ (see, e.g., [14]). On the other hand, a method for fitting BNs with constraint $K = 1$ in polynomial time is available under certain assumptions satisfied in our application. This method is based on the equivalence of the approximation of multivariate probability models using tree-structured dependence and the minimum spanning tree (MST) problem as described in [15]. The objective is the minimization of an information difference $I(P, P_t)$, where $P$ is the target density, and $P_t$ is selected from a class of tree-structured approximating densities. Interest in [15] is restricted to discrete densities. We find, however, that the basic idea extends to general BNs in a natural way. See [16] for further discussion of this model.

Many heuristic or approximate methods exist for fitting Bayesian networks. See [17] for a recent survey. Such algorithms are usually based on MCMC techniques or heuristic algorithms such as TABU searches [18]. We note that the proposed hypothesis test will depend on the calculation of a maximum likelihood ratio, hence it is important to have reasonable guarantees that a maximum has been reached. Thus, given the choice between an exact solution of a restricted class of models or an approximate solution of a general class of models, the former seems preferable. Considering also that in the application described below a solution is required for cases number in "10 s or 100 s" of thousands, a polynomial time exact solution to a restricted class of models appears to be the best choice.

Suppose we are given an $n$-dimensional random vector $X$. We will assume that the density is taken from a parametric family $f^\theta(x) = f^\theta(x_1, \ldots, x_n)$, $\theta \in \Theta$. We write first- and second-order marginal densities $f^{\theta_i}(x_i)$ and $f^{\theta_{ij}}(x_i, x_j)$, with conditional densities $f^{\theta_{ij}}(x_i \mid x_j) = f^{\theta_{ij}}(x_i, x_j)/f^{\theta_j}(x_j)$. For convenience, we introduce a dummy vector component $x_0$, for which $f^{\theta_{i0}}(x_i \mid x_0) = f^{\theta_i}(x_i)$. Let $\mathcal{G}_1$ be the set of DAGs on nodes $(1, \ldots, n)$ with maximum indegree 1. This means that a graph $g \in \mathcal{G}_1$ may be written as a mapping $g : (1, \ldots, n) \to (0, 1, \ldots, n)$. If $i$ has indegree 0 set $g(i) = 0$, otherwise $g(i)$ is the parent node of $i$. We must have $g(i) = 0$ for at least one $i$. For each $g \in \mathcal{G}_1$ let $\Theta^g \subset \Theta$ be the set of parameters admitting the BN decomposition

$$\begin{aligned} f^\theta(x) &= \prod_{i=1}^{n} f^{\theta_{ig(i)}}\left(x_i \mid x_{g(i)}\right) \\ &= \left(\prod_{i=1}^{n} f^{\theta_i}(x_i)\right) \times \left(\prod_{i:g(i)>0} \frac{f^{\theta_{ig(i)}}\left(x_i, x_{g(i)}\right)}{f^{\theta_i}(x_i) f^{\theta_{g(i)}}\left(x_{g(i)}\right)}\right). \end{aligned} \tag{3}$$

Now suppose we are given $N$ independent and complete replicates $\widetilde{X} = (X(1), \ldots, X(N))$ of $X$. Write components

$X(k) = (X_1(k), \ldots, X_n(k))$, $k = 1, \ldots, N$. The log likelihood function becomes, for $\theta \in \Theta^g$,

$$L\left(\theta \mid \widetilde{X}\right) = \sum_{i=1}^{n} L_i(\theta_i) + \sum_{i:g(i)>0} L_{ig(i)}\left(\theta_{ig(i)}\right), \text{where}$$

$$L_i(\theta_i) = \sum_{k=1}^{N} \log\left(f^{\theta_i}(X_i(k))\right), \tag{4}$$

$$L_{ij}\left(\theta_{ij}\right) = \sum_{k=1}^{N} \log\left(\frac{f^{\theta_{ij}}\left(X_i(k), X_j(k)\right)}{f^{\theta_i}(X_i(k)) f^{\theta_j}\left(X_j(k)\right)}\right).$$

Suppose we may construct estimators $\hat{\theta}_i = \hat{\theta}_i(\widetilde{X})$, $\hat{\theta}_{ij} = \hat{\theta}_{ij}(\widetilde{X})$. We then assume there is some selection rule $\hat{\theta}^g = \hat{\theta}^g(\widetilde{X}) \in \Theta^g$ for each $g \in \mathcal{G}_1$. This will typically be the exact or approximate maximum likelihood estimate (MLE) on parameter space $\Theta^g$. We will need the following assumptions.

(A1) For each $g \in \mathcal{G}_1$, $\hat{\theta}_i^g = \hat{\theta}_i$, and $\hat{\theta}_{ig(i)}^g = \hat{\theta}_{ig(i)}$.

(A2) For each $i, j$ we have $L_{ij}(\hat{\theta}_{ij}^g) \geq 0$.

We now consider the problem of maximizing $L^*(g \mid \widetilde{X}) = L(\hat{\theta}^g \mid \widetilde{X})$ over $g \in \mathcal{G}_1$. It will be convenient to isolate the term

$$L_2^*\left(g \mid \widetilde{X}\right) = \sum_{i:g(i)>0} L_{ig(i)}\left(\hat{\theta}_{ig(i)}^g\right). \tag{5}$$

A *spanning tree* on nodes $(1, \ldots, n)$ is an acyclic connected undirected graph. Given edge weights $w_{ij}$, a *minimum spanning tree* (MST) is any spanning tree minimizing the sum of its edge weights among all spanning trees. A number of well-known polynomial time algorithms exist to construct a MST. Two that are commonly described are Prim's and Kruskal's algorithms [19]. Kruskal's algorithm is described in [15]. In the following theorem, the problem of maximizing $L^*(g \mid \widetilde{X})$ is expressed as a MST problem.

**Theorem 1.** *If assumptions* (A1)-(A2) *hold, then maximizing* $L^*(g \mid \widetilde{X})$ *over* $\mathcal{G}_1$ *is equivalent to determining the MST for edge weights* $w_{ij} = -L_{ij}(\hat{\theta}_{ij}^g)$.

*Proof.* Under assumption (A1), from definition (4) it follows that $L^*(g \mid \widetilde{X})$ depends on $g$ only through the term $L_2^*(g \mid \widetilde{X})$. Then suppose $g'$ maximizes $L_2^*(g \mid \widetilde{X})$. For any spanning tree $t$ define $W_t = \sum_{(ij)\in t; i<j} w_{ij}$ and suppose $t'$ minimizes $W_t$. Assume $g'$ is not connected. There must be at least two nodes $i, j$ for which $g(i) = g(j) = 0$, and for which the respective subgraphs containing $i, j$ are unconnected. In this case, extend $g'$ to $g''$ by adding directed edge $(i, j)$. We must have $g'' \in \mathcal{G}_1$, and by (A2) we have $L_2^*(g'' \mid \widetilde{X}) \geq L_2^*(g' \mid \widetilde{X})$. We may therefore assume $g'$ is connected. The undirected graph of $g'$ is a spanning tree, so $W_{t'} \leq -L_2^*(g' \mid \widetilde{X})$.

Next, note that $t'$ can be identified with an element of $\mathcal{G}_1$ by defining any node as a root node, enumerating all paths from the root node to terminal nodes, then assigning edge directions to conform to these paths. This implies $L_2^*(g' \mid \widetilde{X}) \geq -W_{t'}$, which in turn implies $L_2^*(g' \mid \widetilde{X}) = -W_{t'}$, and that $g', t'$ may be selected so that $t'$ can be identified with $g'$. □

*Remark 1.* In general, the optimizing graph from $\mathcal{G}_1$ will not be unique. First, the solution to the MST problem need not be unique. Second, there will always be at least two extensions of a spanning tree to a BN.

Marginal means, variances and, correlations of $X$ are denoted $\mu_i, \sigma_i^2, \rho_{ij}$, leading to parameters $\theta_i = (\mu_i, \sigma_i^2)$, $\theta_{ij} = (\theta_i, \theta_j, \rho_{ij})$. Each parameter in the set $\Theta^g$ represents the class of Gaussian BNs which conform to graph $g$. Following the construction in assumption (A1), let $\hat{\theta}_i = (\overline{X}_i, S_i^2)$, $\hat{\theta}_{ij} = (\hat{\theta}_i, \hat{\theta}_j, R_{ij})$ using summary statistics $\overline{X}_i = N^{-1}\sum_k X_i(k)$, $S_i^2 = N^{-1}\sum_k (X_i(k) - \overline{X}_i)^2$, $R_{ij} = N^{-1}(S_i S_j)^{-1}\sum_k (X_i(k) - \overline{X}_i)(X_j(k) - \overline{X}_j)$. Under the usual parameterization, it can be shown that (omitting constants)

$$L_i\left(\hat{\theta}_i^g\right) = -\left(\frac{N}{2}\right)\log(S_i^2),$$
$$L_{ij}\left(\hat{\theta}_{ij}^g\right) = -\left(\frac{N}{2}\right)\log\left(1 - R_{ij}^2\right), \tag{6}$$

noting that, since $0 \leq R_{ij}^2 \leq 1$, assumption (A2) holds.

## 3. General Maximum Likelihood Ratio Test

Identification of nonhomogeneity between two Bayesian networks will be based on a general maximum likelihood ratio test (MLRT). It is important to note the properties of the MLRT are well understood in parametric inference of limited dimension, and a sampling distribution can be accurately approximated with a large enough sample size. These known properties no longer apply in the type of problem considered here, primarily due to the small sample size, large number of parameters, and the fact that optimization over a discrete space is performed. In addition, the maximum likelihood principle itself favors spurious complexity when no model selection principles are used. While we cannot claim that the MLRT possesses any optimum properties in this application, the use of a permutation procedure will permit accurate estimates of the observed significance level while the use of the restricted model class will control to some degree the degrees of freedom of the model. See, for example, [20] for a general discussion of these issues.

Suppose $\{f_\theta : \theta \in \Theta\}$ is a family of densities defined on some parameter set $\Theta$. We are given two random samples $\widetilde{X} = (X_1, \ldots, X_{n_1})$ and $\widetilde{Y} = (Y_1, \ldots, Y_{n_2})$ from respective densities $f^{\theta_1}$ and $f^{\theta_2}$. Denote pooled sample $\widetilde{XY} = (\widetilde{X}, \widetilde{Y})$. The density of $\widetilde{X}$ and $\widetilde{Y}$, respectively, are $f_{\widetilde{X}}^{\theta_1}(\widetilde{x}) = \prod_{i=1}^{n_1} f^{\theta_1}(x_i)$ and $f_{\widetilde{Y}}^{\theta_2}(\widetilde{y}) = \prod_{i=1}^{n_2} f^{\theta_2}(y_i)$. We consider null hypothesis $H_0 : \theta_1 = \theta_2$. Under $H_0$ the joint density of $\widetilde{XY}$ is $f_{\widetilde{XY}}^{\theta'}(\widetilde{x}, \widetilde{y}) = f_{\widetilde{X}}^{\theta'}(\widetilde{x}) f_{\widetilde{Y}}^{\theta'}(\widetilde{y})$ for some parameter $\theta'$. Assume the existence of maximum likelihood estimators

$\theta_X^* = \arg\max_\theta L(\theta \mid \widetilde{X})$, $\theta_Y^* = \arg\max_\theta L(\theta \mid \widetilde{Y})$, and $\theta_{XY}^* = \arg\max_\theta L(\theta \mid \widetilde{XY})$. The general likelihood ratio statistic in logarithmic scale is then (with large values rejecting $H_0$)

$$\Lambda\left(\widetilde{X}, \widetilde{Y}\right) = L\left(\theta_X^* \mid \widetilde{X}\right) + L\left(\theta_Y^* \mid \widetilde{Y}\right) - L\left(\theta_{XY}^* \mid \widetilde{XY}\right). \quad (7)$$

Asymptotic distribution theory is not relevant here due to small sample size and the fact that optimization is performed in part over a discrete space of models, so a two sample permutation procedure will be used. Permutations will be approximately balanced to reduce spurious variability when a true difference in expression pattern exists (see, e. g., [21] for discussion). This can be done by changing group labels of $\overline{n} \approx n_1 n_2/(n_1 + n_2)$ randomly selecting sample vectors from each of $\widetilde{X}$ and $\widetilde{Y}$. This results in permutation replicate samples $\widetilde{X}^P$ and $\widetilde{Y}^P$. The balanced procedure ensures that each permutation replicate sample contains approximately equal proportions of the original samples.

We now define Algorithm 1.

*Algorithm 1.* (1) Determine $g_1, g_2, g_{12}$ by maximizing $L_2^*(g \mid \widetilde{X})$, $L_2^*(g \mid \widetilde{Y})$, $L_2^*(g \mid \widetilde{X}, \widetilde{Y})$ (MST algorithm).

(2) Set $\Lambda^{\text{obs}} = L^*(g_{12} \mid \widetilde{X}, \widetilde{Y}) - L^*(g_1 \mid \widetilde{X}) - L^*(g_2 \mid \widetilde{Y})$.

(3) Construct $M$ replications $\Lambda_1^P, \ldots, \Lambda_M^P$ in the following way. For each replication $i$, create random replicate samples $\widetilde{X}^P$ and $\widetilde{Y}^P$, then determine $g_1^P, g_2^P$ which maximize $L_2^*(g \mid \widetilde{X}^P)$, $L_2^*(g \mid \widetilde{Y}^P)$. Set $\Lambda_i^P = L^*(g_{12} \mid \widetilde{XY}) - L^*(g_1^P \mid \widetilde{X}^P) - L^*(g_2^P \mid \widetilde{Y}^P)$.

(4) Set $P$-value

$$\widehat{p} = \frac{\left|\left\{\Lambda_i^P \geq \Lambda^{\text{obs}}\right\}\right| + 1}{M + 1}. \quad (8)$$

Note that the quantity $L^*(g_{12} \mid \widetilde{XY})$ is permutation invariant and hence need not be recalculated within the permutation procedure.

## 4. Permutation Tests with Stopping Rules

Permutation or bootstrap tests usually reduce to the estimation of a binomial probability by direct simulation. Since interest is usually in identifying small values, it would seem redundant to continue sampling when, for example, the first ten simulations lead to an estimate of 1/2. This suggests that a stopping rule may be applied to permutation sampling, resulting in significant reduction in computation time, provided it can be incorporated into a valid inference statement. A variety of such procedures have been described in the literature but do not seem to have been widely adopted in genomic discovery applications [22–24].

Suppose, as in Algorithm 1, we have an observed test statistic $\Lambda^{\text{obs}}$, and can simulate indefinitely a sequence $\Lambda_1^P, \Lambda_2^P, \ldots$ from a null distribution $P_0$. By convention we assume that large values of $\Lambda^{\text{obs}}$ tend to reject the null hypothesis. To develop a stopping rule for this sequence set

$$S_i = \sum_{i'=1}^{i} I\left\{\Lambda_{i'}^P \geq \Lambda^{\text{obs}}\right\}. \quad (9)$$

Formally, $T$ is a *stopping time* if the occurrence of event $\{T > t\}$ can be determined from $S_1, \ldots, S_t$. We may then design an algorithm which terminates after sampling a sequence of exactly length $T$ from $P_0$, then outputs $\Lambda_1^P, \ldots, \Lambda_T^P$, from which the hypothesis decision is resolved. We refer to such a procedure as a *stopped procedure*. A *fixed procedure* (such as Algorithm 1) can be regarded as a special case of a stopped procedure in which $T \equiv M$.

An important distinction will have to be made between a single test and a *multiple testing procedure* (MTP), which is a collection of $K$ hypothesis tests with rejection rules that control for a global error rate such as *false discovery rate* (FDR), *family-wise error rate* (FWER), or *per family error rate* (PFER) [25]. In the single test application, we may set a fixed significance level $\alpha$ and continue replications until we conclude that the $P$-value is above or below $\alpha$. For an MTP, it will be important to be able to estimate small $P$-values, so a stopping rule which permits this is needed. Although the two cases have different structure, in our development they will both be based on the *sequential probability ratio test* (SPRT), first proposed in [26], which we now describe.

*4.1. Sequential Probability Ratio Test (SPRT).* Formally (see [27, Chapter 2]) the SPRT tests between two simple alternatives $H_0$: $\theta = \theta_0$ versus $H_1$: $\theta = \theta_1$, where $\theta$ parametrizes a family of distributions $f_\theta$. We assume there is a sequence of *iid* observations $x_1, x_2 \ldots$ from $f_\theta$ where $\theta \in \{\theta_0, \theta_1\}$. Let $l_n(\theta)$ be the likelihood function based on $(x_1, \ldots, x_n)$ and define the likelihood ratio statistic $\lambda_n = l_n(\theta_1)/l_n(\theta_0)$. For two constants $A < 1 < B$, define stopping time

$$T = \min\{n : \lambda_n \notin (A, B)\}. \quad (10)$$

It can be shown that $E_\theta[T] < \infty$. If $\lambda_T \leq A$ we conclude $H_0$ and conclude $H_1$ otherwise. We define errors $\alpha_0 = P_{\theta_0}(\lambda_T \geq B)$ and $\alpha_1 = P_{\theta_1}(\lambda_T \leq A)$. It turns out that the SPRT is optimal under the given assumptions in the sense that it minimizes $E_\theta[T]$ among all sequential tests (which includes fixed sample tests) with respective error probabilities no larger than $\alpha_0, \alpha_1$. Approximate formulae for $\alpha_0, \alpha_1$ and $E_{\theta_0}[T], E_{\theta_1}[T]$ are given in [27].

Hypothesis testing usually involves composite hypotheses, with distinct interpretations for the null and alternative hypothesis. One method of adapting the SPRT to this case is to select surrogate simple hypotheses. For example, to test $H_0 : \theta \geq \theta'$ versus $H_1 : \theta < \theta'$, we could select simple hypotheses $\theta_0 \geq \theta'$ and $\theta_1 < \theta'$. In this case, we would need to know the entire power function, which may be estimated using simulations.

An additional issue then arises in that the expected stopping time may be very large for $\theta \in (\theta_0, \theta_1)$. This can be accommodated using truncation. Suppose a reasonable choice for a fixed sample size is $M$. We would then use truncated stopping time $T^M = \min\{T, M\}$, with $T$ defined in (10). When $T > M$, we could, for example, select hypothesis $H_0$ if $\lambda_M \leq 1$. These modifications are discussed in [27].

*4.2. Single Hypothesis Test.* Suppose we adopt a fixed significance level $\alpha$ for a single hypothesis test. If $\alpha^{\text{obs}}$ is

the (unknown) true significance level, we are interested in resolving the hypothesis $H{:}\alpha^{\text{obs}} \leq \alpha$. The properties of the test are summarized in a power curve, that is, the probability of deciding $H$ is true for each $\alpha^{\text{obs}}$. An example of this procedure is given in [28], for $\alpha = 0.05$, using a SPRT with parameters $A = 0.0010101$, $B = 99.9$, $\theta_0 = 0.03$, $\theta_1 = 0.05$, and truncation at $M = 2000$. Hypothesis $H$ is concluded if $\lambda_{T^M} \leq A$ when $T < M$; otherwise when $\lambda_M \leq 1$.

*4.3. Multiple Hypothesis Tests.* We next assume that we have $K$ hypothesis tests based on sequences of the form (9). We wish to report a global error rate, in which case specific values of small $P$-values are of importance. We will consider specifically the class of MTPs referred to as either *step-up* or *step-down* procedures. If we are given a sequence of $K$ $P$-values $p_1, \ldots, p_K$ which have ranks $\nu_1, \ldots, \nu_K$, then *adjusted P-values*, $p^a_{\nu_i}$ are given by:

$$p^a_{\nu_i} = \max_{j \leq i} \min\left(C\left(K, j, p_{\nu_j}\right), 1\right) \text{(step-down procedure)},$$

$$p^a_{\nu_i} = \min_{j \geq i} \min\left(C\left(K, j, p_{\nu_j}\right), 1\right) \text{(step-up procedure)}, \tag{11}$$

where the quantity $C(K, j, p)$ defines the particular MTP. It is assumed that $C(K, j, p)$ is an increasing function of $p$ for all $K, j$. The procedure is implemented by rejecting all null hypotheses for which $p^a_i \leq \alpha$. Depending on the MTP, various forms of error, usually either *family-wise error rate* (FWER) or *false discovery rate* (FDR), are controlled at the $\alpha$ level. For example, the Benjamini-Hochberg (BH) procedure is a step-up procedure defined by $C(K, j, p) = j^{-1}Kp$ and controls for FDR for independent hypothesis tests. A comprehensive treatment of this topic is given in, for example, [25].

Suppose we have $K$ probabilities $p_1, \ldots, p_K$ ($P$-values associated with $K$ tests). For each test $i = 1, \ldots, K$, we may generate $S^i_j \sim \text{bin}(p_i, j)$ as the cumulative sum defined in (9). Now suppose we define any stopping time $T_i$, bounded by $M$, for each sequence $S^i_1, \ldots, S^i_M$ (this may or may not be related to the SPRT). Then define estimates $\widetilde{p}_i = \widehat{p}_i I\{T_i = M\} + I\{T_i < M\}$, with $\widehat{p}_i = (|\{\Lambda^P_i \geq \Lambda^{\text{obs}}\}| + 1)/(M + 1)$.

For a fixed MTP, the estimates $\widehat{p}_1, \ldots, \widehat{p}_K$ would replace the true values in (11), yielding estimated adjusted $P$-values $\widehat{p}^a_i$ while for the stopped MTP adjusted $P$-values $\widetilde{p}^a_i$ are produced in the same manner using $\widetilde{p}_1, \ldots, \widetilde{p}_K$. It is easily seen that $\widetilde{p}_i \geq \widehat{p}_i$ while the rankings of $\widetilde{p}_i$ (accounting for ties) are equal to the rankings of $\widehat{p}_i$. Furthermore, the formulae in (11) are monotone in $p_i$, so we must have $\widetilde{p}^a_i \geq \widehat{p}^a_i$. Thus, the stopped procedure may be seen as being embedded in the fixed procedure. It inherits whatever error control is given for the fixed MTP, with the advantage that the calculation of the adjusted $P$-values $\widetilde{p}^a_i$ uses only the first $T_i$ replications for the $i$th test.

The procedure will always be correct in that it is strictly more conservative than the fixed MTP in which it is embedded, no matter which stopping time is used. The remaining issue is the selection of $T_i$ which will equal $M$ for small enough values of $p_i$ but will also have $E[T_i] \ll M$ for larger values of $p_i$. It is a simple matter, then, to modify the SPRT described in Section 4.2 by eliminating the lower bound $A$ (equivalently $A = 0$). We will adopt this design in this paper. This gives Algorithm 2.

*Algorithm 2.* (1) Same as Algorithm 1, step 1.

(2) Same as Algorithm 1, step 2.

(3) Simulate replicates $\Lambda^P_i$ in Algorithm 1, step 3, until the following stopping criterion is met. Set $S_i = \sum_{i'=1}^{i} I\{\Lambda^P_{i'} \geq \Lambda^{\text{obs}}\}|$, and let $\lambda_i = [\theta_1/\theta_0]^{S_i}[(1 - \theta_1)/1 - \theta_0]^{i-S_i}$, where $\theta_0 \leq \alpha < \theta_1$. Stop sampling at the $i$th replication if $\lambda_i \geq B$, where $B > 1$, or until $i = M$, whichever occurs first.

(4) Let $T'$ be the number of replications in step 3. If $T' = M$, set

$$\widetilde{p} = \frac{\left|\left\{\Lambda^P_i \geq \Lambda^{\text{obs}}\right\}\right| + 1}{M + 1}, \tag{12}$$

otherwise set $\widetilde{p} = 1$.

The values $\widetilde{p}$ generated by Algorithm 2 can then be used in a stopped MTP as described in this section.

## 5. Gene-Set Analysis

A recent trend in the analysis of microarray data has been to base the discovery of phenotype-induced DE on gene sets rather than individual genes. The reasoning is that if genes in a given set are related by common pathway membership or other transcriptional process, then there should be an aggregate change in gene expression pattern. This should give increased statistical power, as well as enhanced interpretability, especially given the lack of reproducibility in univariate gene discovery due to the stringent requirements imposed by multiple testing adjustments. Thus, the discovery process reduces to a much smaller number of hypothesis tests with more direct biological meaning. Some objections may be raised concerning the selection of the gene sets when theses sets are themselves determined experimentally. Additionally, gene sets may overlap. While these problems need to be addressed, it is also true that such gene set methods have been shown to detect DE not uncovered by univariate screens.

A crucial problem in gene set analysis is the choice of test statistic. The problem of testing against equality of random vectors in $\Re^d$, $d > 1$, is fundamentally different from the univariate case $d = 1$. The range of statistics one would consider for $d = 1$ is reasonably limited, the choice being largely driven by distributional considerations. For $d > 1$, new structural or geometric considerations arise. For example, we may have differential expression between some but not all genes in the gene set, which makes selection of a single optimal test statistic impossible. Alternatively, the experimental random vectors may differ in their level of coexpression independently of their level of marginal DE.

In fact, almost all GS procedures directly measure aggregate DE, so an important question is whether or not phenotypic variation is almost completely expressible

as DE. If so, then a DE based statistic will have fewer degrees of freedom, hence more power, than one based on a more complex model. Otherwise, a reasonable conjecture is that a compound GS analysis will work best, employing a DE statistic as well as one more sensitive to changes in coexpression patterns.

Correlations have been used in a number of gene discovery applications. They may be used to associate genes of unknown function with known pathways [29, 30]. Additionally, a number of GS procedures exist which incorporate correlation structure into the procedure [31–33]. However, a direct comparison of correlations is not practical due to the large number $(d(d-1)/2)$ of distinct correlation parameters. Therefore, there is a considerable advantage to the statistic (7) based on the reduced BN model, in that the correlation structure can be summarized by the $d$ correlation parameters output by the MST algorithm, yielding a transitive dependence model similar to that effectively exploited in [29].

It is important to refer to a methodological characterization given in [34]. A distinction is made between two types of null hypotheses. Suppose we are given samples of expression levels from a gene set $G$ from two phenotypes. Suppose also that for each gene in $G$ and its complement $G^c$, a statistical measure of differential expression is available. For a *competitive test*, the null hypothesis $H_0^{\text{comp}}$ is that the prevalence of differential expression in $G$ is no greater than in $G^c$. For a *self-contained test*, the null hypothesis $H_0^{\text{self}}$ is that no genes in $G$ are differentially expressed. In the GSEA method of [4, 5] concern is with $H_0^{\text{comp}}$. In most subsequent methods, including the one proposed here, $H_0^{\text{self}}$ is used.

For general discussions of the issues raised here, see [35–37]. Comprehensive surveys of specific methods can be found in [38] or [39].

*5.1. Experimental Data.* We will demonstrate the algorithm proposed here on two data sets examined elsewhere in the literature. These were obtained from the GSEA website *www.broad.mit.edu/gsea* [6]. In [5], a data set *p53* is extracted from the NCI-60 collection of cancer cell lines, with 17 cell lines classified as normal, and 33 classified as carrying mutations of p53. We also examine the *DIABETES* data set introduced in [4], consisting of microaray profiles of skeletal muscle biopsies from 43 males. For the *DIABETES* data set used here, there were 17 normal glucose tolerance (*NGT*) subjects and 17 diabetes (*DMT*) subjects. For gene sets, we used one of the gene set lists compiled in [5], denoted $C_2$, consisting of 472 gene sets with products collectively involved in various metabolic and signalling pathways, as well as 50 sets containing genes exhibiting coregulated response to various perturbations. In our analyses, FDR will be estimated using the BH procedure.

*5.1.1. P53 Data.* A $t$-test was performed on each of the 10,100 genes. Only 1 gene had an adjusted $P$-value less than FDR = 0.25 (*bax*, $P = 5 \times 10^{-6}$, $P_{\text{adj}} = 0.05$). Several GS analyses for this data set (using $C_2$) have been reported. We cite the GSEA analysis in [5] and a modification of the GSEA proposed in [40]. Also, in [38], this data set is used to test three procedures, each using various standardization procedures. Two are based on logistic regression (*Global test* [41] *ANCOVA Global test* [42]). The third is an extension of the *Significance Analysis of Microarray* (SAM) procedure [43] to gene sets proposed in [44] (SAM-GS).

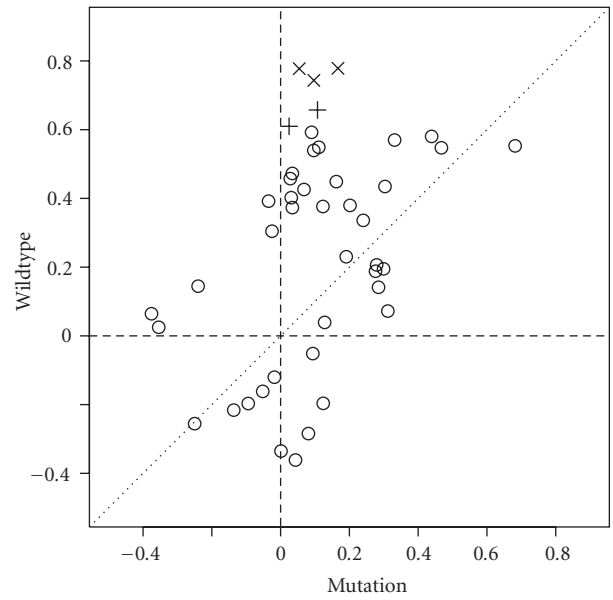


Figure 1: Scatterplot of correlations for all gene pairs in cell_cycle_checkpoint_II pathway, using wildtype and mutation axes. Genes with nominal significance levels for differential coexpression $P \in (.01, .05]$ (×) and $P \leq .01$ (+) are indicated separately.

Table 1 lists pathways selected from $C_2$ for the analysis proposed here using FDR ≤ 0.25, including unadjusted and adjusted $P$-values. For each entry we indicate whether or not the pathway was selected under the analyses reported in [5] (*Sub*, FDR ≤ 0.25), [40] (*Efr*, FDR ≤ 0.1) and [38] (*Liu*, nominal $P$-value ≤ .001 in at least one procedure). It is important to note that the results indicated with an asterisk (*) are not directly comparable due to differing MTP control, and are included for completeness.

The first five pathways are directly comparable. Of these, two were not detected in any other analysis. Our procedure was repeated for these pathways using the sum of the squared t-statistics across genes. The nominal $P$-values for *g2 Pathway* and *cell cycle checkpoint II* were.0044 and >.05, respectively. Since we are interested in identifying pathways which may be detectable by pathway methods, but not DE based methods we will examine *cell cycle checkpoint II* more closely. Applying a univariate $t$-test to each of the 10 genes yields one $P$-value of 0.001 (*cdkn2a*), with the remaining $P$-values greater than 0.1 hence a DE-based approach is unlikely to select this pathway. Furthermore, $P$-values under 0.05 for change in correlation are reported for *rbbp8/rb1*, *nbs1/ccng2*, *atr/ccne2*, *nbs1/tp53, and ccng2/tb53* ($P$ = .002, .006, .008, .035, and .036). Clearly, the difference in gene expression pattern is determined by change in coexpression pattern. In Figure 1, the correlations for all gene pairs for wild-type and mutation

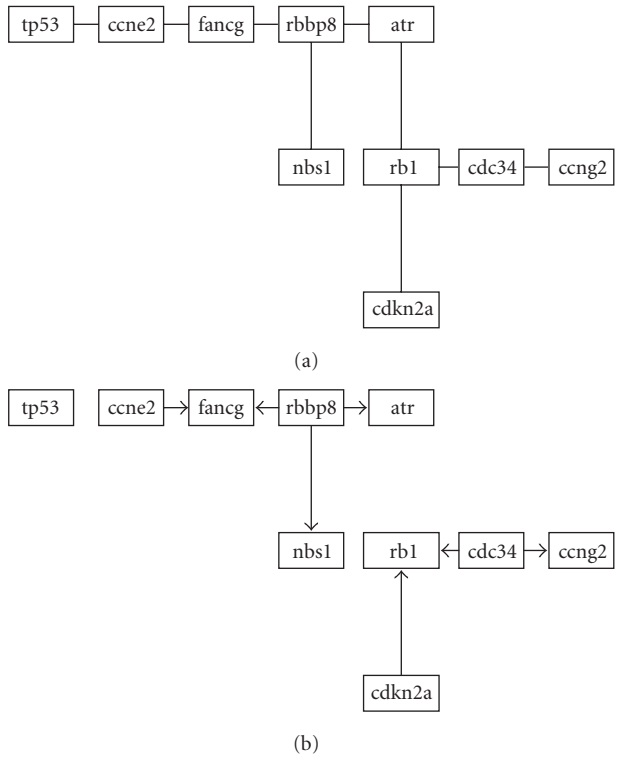


Figure 2: Bayesian network fits for mutation data for cycle checkpoint II pathway using (a) Minimum Spanning Tree algorithm (maximum indegree of 1); (b) Bayesian Information Criterion (maximum indegree of 2).

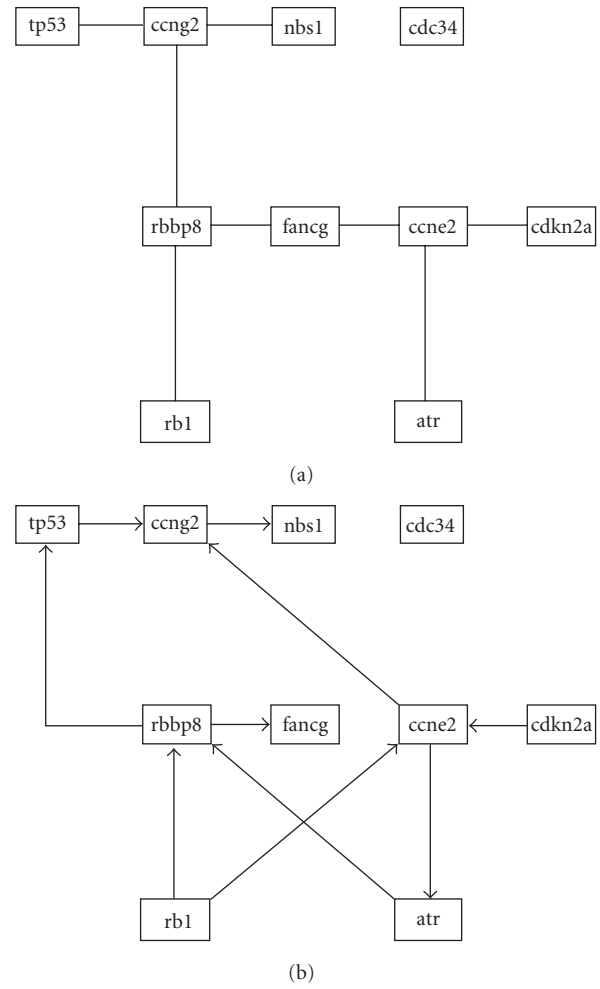


Figure 3: Bayesian network fits for wildtype data for cycle checkpoint II pathway using (a) Minimum Spanning Tree algorithm (maximum indegree of 1). (b) Bayesian Information Criterion (maximum indegree of 2).

groups are indicated. A clear pattern is evident, by which correlation structure present in the wildtype class does not exist in the mutation class.

To further clarify the procedure, we compare the BN model obtained from the data for the ten genes associated with the *cell cycle checkpoint II* pathway, separately for mutation and wildtype conditions. If there is interest in a post-hoc analysis of any particular pathway, the rational for the MST algorithm no longer holds, since only one fit is required. It is therefore instructive to compare the MST model to a more commonly used method. In this case, we will use the Bayesian Information Criterion (BIC) (see, e.g., [7]), with a maximum indegree of 2. To fit the model we use a simulated annealing algorithm adapted from [45]. The resulting graphs are shown in Figures 2 (mutation) and 3 (wildtype). The MST and BIC fits are labelled (a) and (b) respectively. For the mutation fit, there is a very close correspondence between the topologies produced by the respective methods. For the wildtype data, some correspondence still exists, but less so then for the mutation data. The topologies between the conditions differ more significantly, as predicted by the hypothesis test.

*5.1.2. Diabetes.* No pathways were detected at a FDR of 0.25. The two pathways with the smallest $P$-values were *atrbrca Pathway* and *MAP00252 Alanine and aspartate metabolism* ($P = .0026, .003$). In [33] the latter pathway was the single pathway reported with PFER = 1. The comparable PFER rate of the two pathways reported here would be 1.36 and 1.57. The *atrbrca Pathway* contains 25 genes. Of these, only *fance* differentially expressed at a 0.05 significance level ($P = .0059$). For each gene pair, correlation coefficients were calculated and tested for equality between classes *NGT* and *DMT*. Table 2 lists the 10 highest ranking gene pairs in terms of correlation magnitude within the *NGT* class. Also listed is the corresponding correlation within the *DMT* class, as well as the two-sample $P$-value for correlation difference. The analysis is repeated after exchanging classes, also in Table 2. We note that for a sample size of 17, an approximate 95% confidence interval for a reported correlation of $R = 0.6$ is $(0.17, 0.84)$ whereas the standard deviation of a sample correlation coefficient of mean zero is approximately 0.27. There is likely to be considerable statistical variation in graphical structure under the null hypothesis.

Examining the first table, differences in correlation appear to be explainable by sampling variation. In the second there are two gene pairs *fanca/fance* and *fanca/hus1* with

Table 1: P53 pathways, with GS size ($N$), unadjusted and FDR adjusted $P$-values ($P, P^a$). Inclusion in analyses cited in Section 5.1 indicated. †The complete name of DNA_DAMAGE is DNA_DAMAGE_SIGNALLING. ‡The complete name of MAP00562 is MAP00562_Inositol_phosphate_metabolism. *Inclusion criterion based on control rate of original analysis.

| Pathway | $N$ | $P$ | $P^a$ | Sub | Efr | Liu |
|---|---|---|---|---|---|---|
| SA_G1_AND_S_PHASES | 14 | <.001 | .08 | n | y | n |
| atmPathway | 19 | <.001 | .08 | n | n | y |
| g2Pathway | 23 | <.001 | .08 | n | n | n |
| p53Pathway | 16 | <.001 | .08 | y | y | y |
| cell_cycle_checkpointII | 10 | <.001 | .08 | n | n | n |
| SA_FAS_SIGNALLING | 9 | .002 | .14 | n | n* | n* |
| cellcyclePathway | 23 | .002 | .16 | n | n* | n* |
| DNA_ DAMAGE† | 90 | .003 | .17 | n | n* | n* |
| SA_TRKA_RECEPTOR | 16 | .003 | .17 | n | n* | y* |
| radiation_sensitivity | 26 | .003 | .17 | y | y* | y* |
| ngfPathway | 19 | .004 | .17 | n | y* | n* |
| GO_ROS | 23 | .004 | .17 | n | n* | n* |
| etsPathway | 16 | .004 | .17 | n | n* | n* |
| ck1Pathway | 15 | .006 | .21 | n | n* | n* |
| erkPathway | 29 | .007 | .23 | n | n* | n* |
| MAP00562‡ | 18 | .007 | .23 | n | n* | n* |
| arfPathway | 13 | .007 | .23 | n | n* | n* |

Table 2: Correlation analysis for *DIABETES* data. For each pathway and phenotype, 10 gene pairs with the largest correlation (×100) magnitudes; correlation (×100) of alternative phenotype; and $P$-value (×1000) against equality.

| atr brca pathway | | | | Alanine pathway | | | |
|---|---|---|---|---|---|---|---|
| *NGT* | | cor | | *NGT* | | cor | |
| genes | *ngt* | *dmt* | $P$ | genes | *ngt* | *dmt* | $P$ |
| fancc/rad17 | 83 | 69 | 349 | crat/got1 | 81 | 30 | 031 |
| fancc/brca2 | 76 | 44 | 156 | nars/dars | 80 | −24 | <1 |
| rad9a/rad17 | 76 | 87 | 338 | crat/gpt | 75 | 15 | 028 |
| chek2/rad17 | 71 | 35 | 172 | got2/adss | −75 | −02 | 012 |
| brca1/hus1 | −69 | −29 | 148 | got2/abat | −73 | 34 | 001 |
| rad17/brca2 | 67 | 56 | 632 | ddx3x/got1 | 72 | −17 | 004 |
| atr/mre11a | −64 | −41 | 403 | crat/ass | 72 | 12 | 037 |
| chek1/nbs1 | −62 | 09 | 030 | ddx3x/dars | 71 | 12 | 043 |
| rad51/rad1 | −62 | −23 | 198 | gpt/got1 | 70 | 33 | 175 |
| rad9a/fancc | 59 | 76 | 388 | ddx3x/abat | −68 | −41 | 305 |
| *DMT* | | cor | | *DMT* | | cor | |
| genes | *dmt* | *ngt* | $P$ | genes | *dmt* | *ngt* | $P$ |
| rad9a/rad17 | 87 | 76 | 338 | ddx3x/aars | −76 | −55 | 325 |
| fanca/fance | 81 | 14 | 009 | crat/nars | 74 | 26 | 074 |
| rad9a/fancc | 76 | 59 | 388 | ddx3x/nars | 73 | 66 | 715 |
| fanca/hus1 | −72 | 27 | 002 | asns/ddo | 60 | 42 | 502 |
| brca1/mre11a | 71 | 11 | 039 | pc/aars | −58 | 15 | 031 |
| fancc/rad17 | 69 | 83 | 349 | crat/pc | 58 | 53 | 862 |
| fancf/hus1 | 67 | 53 | 563 | crat/ddx3x | 58 | 51 | 813 |
| brca1/atr | −67 | 16 | 011 | got1/dars | −56 | 40 | 006 |
| rad17/mre11a | 64 | 11 | 086 | pc/nars | 55 | 18 | 244 |
| fancg/rad51 | 64 | 22 | 160 | asns/gad2 | −54 | −44 | 723 |

Table 3: For stopped (*St*) and fixed (*Fx*) procedures, the table gives computation times; mean number of replications; % gene sets completely sampled; number of pathways with *P*-values ≤.01; and number of such pathways in agreement.

| Data | Time (*hrs*) | | Mean rep | | % comp | | $^{\#}P \leq .01$ | | |
|---|---|---|---|---|---|---|---|---|---|
| | *St* | *Fx* | *St* | *Fx* | *St* | *Fx* | *St* | *Fx* | Both |
| *diab* | 3.7 | 35.8 | 341.0 | 5000 | 5.4 | 100 | 6 | 6 | 6 |
| *p53* | 2.1 | 30.0 | 612.3 | 5000 | 10.5 | 100 | 18 | 19 | 18 |

small *P*-values (.009, .002). We note that they share a common gene *fanca* and that they involve the only gene *fance* exhibiting differential expression. The correlation patterns within the two samples are otherwise similar, suggesting a specific alteration of the network model.

The situation differs for the pathway *MAP00252 Alanine and aspartate metabolism*, summarized in Table 2 using the same analysis. The change in correlation is more widespread. The 8 gene pairs with the highest correlation magnitudes within the *NGT* sample differ between *NGT* and *DMT* at a 0.05 significance level. Furthermore, the number of gene pairs with correlation magnitudes exceeding 0.7 is 9 in the *NGT* sample, but only 3 in the *DMT* sample.

*5.1.3. Comparison of Fixed and Stopped Procedures.* Both the fixed and stopped procedures were applied to the preceding analysis. The SPRT used parameters $A = 0$, $B = 99.9$, $\theta_0 = 0.05$, $\theta_1 = 0.07$, and truncation at $M = 5000$. Table 3 summarizes the computation times for each method as well as the selection agreement. In these examples, the stopped procedure required significantly less computation time with no apparent loss in power.

## 6. Conclusion

We have introduced a two-sample general likelihood ratio test for the equality of Bayesian network models. Significance levels are estimated using a permutation procedure. The algorithm was proposed as an alternative form of gene-set analysis. It was noted that the fitting of Bayesian networks is computationally time consuming, hence a need for the efficient calculation of a model fit was identified, particularly for this application.

Two procedures were introduced to meet this requirement. First, we implemented a version of a minimum spanning tree algorithm first proposed in [15] which permits the polynomial-time calculation of the maximum likelihood Bayesian network among those with maximum indegree of one. Second, we introduced sequential testing principles to the problem of multiple testing, finding that a straightforward stopping rule could be developed which preserves group error rates for a wide range of procedures.

We may expect this form of test to be especially sensitive to changes in coexpression patterns, in contrast to most gene-set procedures, which directly measure aggregate differential expression. In an application of the algorithm to two data sets considered in [5], a number of selected gene-sets exhibited clear differences in coexpression patterns while exhibiting very little differential expression. This leads to the conjecture that the optimal approach to gene-set analysis is to couple a test which directly measures aggregate differential expression with one designed to detect differential coexpression.

## Acknowledgments

This paper was supported by NIH Grant no. R21HG004648. The Clinical Translational Science Institute of the University of Rochester Medical Center also provided funding for this research.

## References

[1] E. R Dougherty, I. Shmulevich, J. Chen, and Z. J. Wang, *Genomic Signal Processing and Statistics*, vol. 2 of *EURASIP Book Series on Signal Processing and Communications*, Hindawi Publishing Corporation, New York, NY, USA, 2005.

[2] I. Shmulevich and E. R. Dougherty, *Genomic Signal Processing*, Princeton University Press, Princeton, NJ, USA, 2007.

[3] F. Emmert-Streib and M. Dehmer, "Detecting pathological pathways of a complex disease by a comparitive analysis of networks," in *Analysis of Microarray Data: A Network-Based Approach*, F. Emmert-Streib and M. Dehmer, Eds., pp. 285–305, Wiley-VCH, Weinheim, Germany, 2008.

[4] V. K. Mootha, C. M. Lindgren, K.-F. Eriksson et al., "PGC-1$\alpha$-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes," *Nature Genetics*, vol. 34, no. 3, pp. 267–273, 2003.

[5] A. Subramanian, P. Tamayo, V. K. Mootha et al., "Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 43, pp. 15545–15550, 2005.

[6] A. Subramanian, H. Kuehn, J. Gould, P. Tamayo, and J. P. Mesirov, "GSEA-P: a desktop application for gene set enrichment analysis," *Bioinformatics*, vol. 23, no. 23, pp. 3251–3253, 2007.

[7] P. Sebastiani, M. Abad, and M. F. Ramoni, "Bayesian networks for genomic analysis," in *Genomic Signal Processing and Statistics*, E. R. Dougherty, I. Shmulevich, J. Chen, and Z. J. Wang, Eds., EURASIP Book Series on Signal Processing and Communications, Hindawi Publishing Corporation, New York, NY, USA, 2005.

[8] N. Friedman, M. Linial, I. Nachman, and D. Pe'er, "Using Bayesian networks to analyze expression data," *Journal of Computational Biology*, vol. 7, no. 3-4, pp. 601–620, 2000.

[9] C. J. Needham, J. R. Bradford, A. J. Bulpitt, and D. R. Westhead, "A primer on learning in Bayesian networks for computational biology.," *PLoS Computational Biology*, vol. 3, no. 8, p. e129, 2007.

[10] T. Chu, C. Glymour, R. Scheines, and P. Spirtes, "A statistical problem for inference to regulatory structure from associations of gene expression measurements with microarrays," *Bioinformatics*, vol. 19, no. 9, pp. 1147–1152, 2003.

[11] R. G. Cowell, P. Dawid, S. L. Lauritzen, and D. J. Spiegelhalter, *Probabilistic Networks and Expert Systems: Exact Computational Methods for Bayesian Networks*, Information Science and Statistics, Spring, New York, NY, USA, 1999.

[12] R. G. Cowell, "Efficient maximum likelihood pedigree reconstruction," *Theoretical Population Biology*, vol. 76, no. 4, pp. 285–291, 2009.

[13] T. Silander and P. Myllymki, "A simple approach to finding the globally optimal bayesian network structure," in *Proceedings of the 22nd Conference on Artificial intelligence (UAI '06)*, R. Dechter and T. Richardson, Eds., pp. 445–452, AUAI Press, 2006.

[14] D. M. Chickering, "Learning Bayesian net- works is NP-complete," in *Learning from Data: Artificial Intelligence and Statistics V*, D. Fisher and H. Lenz, Eds., pp. 121–130, Springer, New York, NY, USA, 1996.

[15] C. K. Chow and C. N. Liu, "Approximating discrete probability distributions with dependence trees," *IEEE Transactions on Information Theory*, vol. 14, pp. 462–467, 1968.

[16] P. Abbeel, D. Koller, and A. Y. Ng, "Learning factor graphs in polynomial time and sample complexity," *Journal of Machine Learning Research*, vol. 7, pp. 1743–1788, 2006.

[17] K. Murphy, "Software packages for graphical models bayesian networks," *Bulletin of the International Society for Bayesian Analysis*, vol. 14, pp. 13–15, 2007.

[18] M. Teyssier and D. Koller, "Ordering-based search: a simple and effective algorithm for learning bayesian networks," in *Proceedings of the 21st Conference on Uncertainty in AI (UAI '05)*, pp. 584–590, 2005.

[19] C. H. Papadimitriou and K. Steiglitz, *Combinatorial Optimization: Algorithms and Complexity*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1982.

[20] A. H. Walsh, *Aspects of Statistical Inference*, John Wiley & Sons, New York, NY, USA, 1996.

[21] B. Efron, "Robbins, empirical Bayes and microarrays," *Annals of Statistics*, vol. 31, no. 2, pp. 366–378, 2003.

[22] J. Besag and P. Clifford, "Sequential monte carlo $p$-values," *Biometrika*, vol. 78, pp. 301–304, 1991.

[23] R. H. Lock, "A sequential approximation to a permutation test," *Communications in Statistics. Simulation and Computation*, vol. 20, no. 1, pp. 341–363, 1991.

[24] M. P. Fay and D. A. Follmann, "Designing Monte Carlo implementations of permutation or bootstrap hypothesis tests," *American Statistician*, vol. 56, no. 1, pp. 63–70, 2002.

[25] S. Dudoit and M. J. van der Laan, *Multiple Testing Procedures with Applications to Genomics*, Springer, New York, NY, USA, 2008.

[26] A. Wald, *Sequential Analysis*, John Wiley & Sons, New York, NY, USA, 1947.

[27] D. Siegmund, *Sequential Analysis: Tests and Confidence Intervals*, Springer, New York, NY, USA, 1985.

[28] A. Almudevar, "Exact confidence regions for species assignment based on DNA markers," *Canadian Journal of Statistics*, vol. 28, no. 1, pp. 81–95, 2000.

[29] X. Zhou, M.-C.J. Kao, and W. H. Wong, "Transitive functional annotation by shortest-path analysis of gene expression data," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 20, pp. 12783–12788, 2002.

[30] R. Braun, L. Cope, and G. Parmigiani, "Identifying differential correlation in gene/pathway combinations," *BMC Bioinformatics*, vol. 9, article no. 488, 2008.

[31] W. T. Barry, A. B. Nobel, and F. A. Wright, "Significance analysis of functional categories in gene expression studies: a structured permutation approach," *Bioinformatics*, vol. 21, no. 9, pp. 1943–1949, 2005.

[32] Z. Jiang and R. Gentleman, "Extensions to gene set enrichment," *Bioinformatics*, vol. 23, no. 3, pp. 306–313, 2007.

[33] L. Klebanov, G. Glazko, P. Salzman, A. Yakovlev, and Y. Xiao, "A multivariate extension of the gene set enrichment analysis," *Journal of Bioinformatics and Computational Biology*, vol. 5, no. 5, pp. 1139–1153, 2007.

[34] J. J. Goeman and P. Bühlmann, "Analyzing gene expression data in terms of gene sets: methodological issues," *Bioinformatics*, vol. 23, no. 8, pp. 980–987, 2007.

[35] D. B. Allison, X. Cui, G. P. Page, and M. Sabripour, "Microarray data analysis: from disarray to consolidation and consensus," *Nature Reviews Genetics*, vol. 7, no. 1, pp. 55–65, 2006.

[36] A. Bild and P. G. Febbo, "Application of a priori established gene sets to discover biologically important differential expression in microarray data," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 43, pp. 15278–15279, 2005.

[37] T. Manoli, N. Gretz, H.-J. Gröne, M. Kenzelmann, R. Eils, and B. Brors, "Group testing for pathway analysis improves comparability of different microarray datasets," *Bioinformatics*, vol. 22, no. 20, pp. 2500–2506, 2006.

[38] Q. Liu, I. Dinu, A. J. Adewale, J. D. Potter, and Y. Yasui, "Comparative evaluation of gene-set analysis methods," *BMC Bioinformatics*, vol. 8, article no. 431, 2007.

[39] M. Ackermann and K. Strimmer, "A general modular framework for gene set enrichment analysis," *BMC Bioinformatics*, vol. 10, article no. 47, 2009.

[40] B. Efron and R. Tibshirani, "On testing the significance of sets of genes," *Annals of Applied Statistics*, vol. 1, pp. 107–129, 2007.

[41] J. J. Goeman, S. van de Geer, F. de Kort, and H. C. van Houwellingen, "A global test for groups fo genes: testing association with a clinical outcome," *Bioinformatics*, vol. 20, no. 1, pp. 93–99, 2004.

[42] U. Mansmann and R. Meister, "Testing differential gene expression in functional groups: goeman's global test versus an ANCOVA approach," *Methods of Information in Medicine*, vol. 44, no. 3, pp. 449–453, 2005.

[43] V. G. Tusher, R. Tibshirani, and G. Chu, "Significance analysis of microarrays applied to the ionizing radiation response," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 9, pp. 5116–5121, 2001.

[44] I. Dinu, J. D. Potter, T. Mueller et al., "Improving gene set analysis of microarray data by SAM-GS," *BMC Bioinformatics*, vol. 8, article 242, 2007.

[45] A. Almudevar, "A simulated annealing algorithm for maximum likelihood pedigree reconstruction," *Theoretical Population Biology*, vol. 63, no. 2, pp. 63–75, 2003.


## Preliminary call for papers

The 2011 European Signal Processing Conference (EUSIPCO-2011) is the nineteenth in a series of conferences promoted by the European Association for Signal Processing (EURASIP, www.eurasip.org). This year edition will take place in Barcelona, capital city of Catalonia (Spain), and will be jointly organized by the Centre Tecnològic de Telecomunicacions de Catalunya (CTTC) and the Universitat Politècnica de Catalunya (UPC).

EUSIPCO-2011 will focus on key aspects of signal processing theory and applications as listed below. Acceptance of submissions will be based on quality, relevance and originality. Accepted papers will be published in the EUSIPCO proceedings and presented during the conference. Paper submissions, proposals for tutorials and proposals for special sessions are invited in, but not limited to, the following areas of interest.

## Areas of Interest

- Audio and electro-acoustics.
- Design, implementation, and applications of signal processing systems.
- Multimedia signal processing and coding.
- Image and multidimensional signal processing.
- Signal detection and estimation.
- Sensor array and multi-channel signal processing.
- Sensor fusion in networked systems.
- Signal processing for communications.
- Medical imaging and image analysis.
- Non-stationary, non-linear and non-Gaussian signal processing.

## Submissions

Procedures to submit a paper and proposals for special sessions and tutorials will be detailed at www.eusipco2011.org. Submitted papers must be camera-ready, no more than 5 pages long, and conforming to the standard specified on the EUSIPCO 2011 web site. First authors who are registered students can participate in the best student paper competition.

## Important Deadlines:

| | |
|---|---|
| Proposals for special sessions | 15 Dec 2010 |
| Proposals for tutorials | 18 Feb 2011 |
| **Electronic submission of full papers** | **21 Feb 2011** |
| Notification of acceptance | 23 May 2011 |
| Submission of camera-ready papers | 6 Jun 2011 |

**Webpage: www.eusipco2011.org**

**Organizing Committee**

***Honorary Chair***
Miguel A. Lagunas (CTTC)

***General Chair***
Ana I. Pérez-Neira (UPC)

***General Vice-Chair***
Carles Antón-Haro (CTTC)

***Technical Program Chair***
Xavier Mestre (CTTC)

***Technical Program Co-Chairs***
Javier Hernando (UPC)
Montserrat Pardàs (UPC)

***Plenary Talks***
Ferran Marqués (UPC)
Yonina Eldar (Technion)

***Special Sessions***
Ignacio Santamaría (Unversidad de Cantabria)
Mats Bengtsson (KTH)

***Finances***
Montserrat Nájar (UPC)

***Tutorials***
Daniel P. Palomar (Hong Kong UST)
Beatrice Pesquet-Popescu (ENST)

***Publicity***
Stephan Pfletschinger (CTTC)
Mònica Navarro (CTTC)

***Publications***
Antonio Pascual (UPC)
Carles Fernández (CTTC)

***Industrial Liaison & Exhibits***
Angeliki Alexiou (University of Piraeus)
Albert Sitjà (CTTC)

***International Liaison***
Ju Liu (Shandong University-China)
Jinhong Yuan (UNSW-Australia)
Tamas Sziranyi (SZTAKI -Hungary)
Rich Stern (CMU-USA)
Ricardo L. de Queiroz (UNB-Brazil)